

# New Optimization Method for Conformational Energy Calculations on Polypeptides: Conformational Space Annealing

JOOYOUNG LEE,<sup>1,2</sup> HAROLD A. SCHERAGA,<sup>2</sup> S. RACKOVSKY<sup>1</sup>

<sup>1</sup>Department of Biomathematical Sciences, Box 1023, Mt. Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029 <sup>2</sup>Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

Received 12 September 1996; accepted 8 January 1997

**ABSTRACT:** A new optimization method is presented to search for the global minimum-energy conformations of polypeptides. The method combines essential aspects of the build-up procedure and the genetic algorithm, and it introduces the important concept of "conformational space annealing." Instead of considering a single conformation, attention is focused on a population of conformations while new conformations are obtained by modifying a "seed conformation." The annealing is carried out by introducing a distance cutoff,  $D_{\text{cut}}$ , which is defined in the conformational space;  $D_{\text{cut}}$  effectively divides the whole conformational space of local minima into subdivisions. The value of  $D_{\text{cut}}$  is set to a large number at the beginning of the algorithm to cover the whole conformational space, and annealing is achieved by slowly reducing it. Many distinct local minima designed to be distributed as far apart as possible in conformational space are investigated simultaneously. Therefore, the new method finds not only the global minimum-energy conformation but also many other distinct local minima as by-products. The method is tested on Met-enkephalin, a 24-dihedral angle problem. For all 100 independent runs, the accepted global minimum-energy conformation was obtained after about 2600 minimizations on average. © 1997 by John Wiley & Sons, Inc. *J Comput Chem* 18: 1222–1232, 1997

**Keywords:** optimization method; global minimum-energy conformations; polypeptides; conformational space annealing

Correspondence to: S. Rackovsky; e-mail: shelly@msvax.mssm.edu

Contract/grant sponsor: National Institute on Aging; contract/grant number: AG-00322

Contract/grant sponsor: National Institute of General Medical Sciences of the National Institutes of Health; contract/grant number: GM-14312

## Introduction

Obtaining global minimum-energy conformations of polypeptides is a very hard optimization problem. The difficulty generally arises from the following two factors: the conformational space of a reasonable size polypeptide is very large, and there are many local minima that are hard to sample efficiently. The energy landscape in the conformational space is very rugged, and there are many large barriers between local minima.

Among many optimization methods aimed at solving this hard optimization problem, we will discuss three methods as a prelude to a new one developed here: the buildup procedure, the genetic algorithm, and the Monte Carlo minimization (MCM) method.

## BUILDUP PROCEDURE

In this method one starts with the low-energy structures of single residues and uses these to build up low-energy structures of dipeptides, tripeptides, and so on, carrying out energy minimization at each level.<sup>1,2</sup> The general idea is to break down the original peptide into several levels of component peptides whose low-energy conformations can be identified by exhaustive enumeration. First, at the single-residue level, all local-minimum conformations are obtained by a grid search of the dihedral angles of each residue of the original peptide and are subsequently stored. At the dipeptide level the stored local-minimum conformations of two single residues are combined and their energies are minimized to obtain local-minimum conformations of dipeptides, which are again stored in order to generate larger oligopeptides, and so on. The problem with this approach is that the number of local minima increases exponentially with the size of the original peptide. To alleviate this difficulty, various strategies, including elimination of high-energy conformations of the component peptides, were used to reduce the number enumerated. In addition, overlapping of peptides was used to reduce redundancy; for example, the tetrapeptide sequence A-B-C-D was built from the tripeptides A-B-C and B-C-D by eliminating all conformations in the tripeptide ensembles A-B-C and B-C-D in which residues B-C of both ensembles were not in the same conformational space.<sup>1</sup>

## GENETIC ALGORITHM

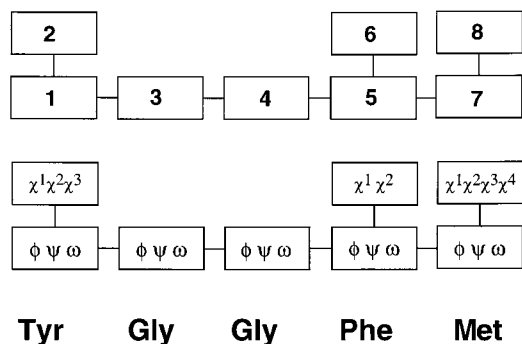
The genetic algorithm<sup>3,4</sup> is based on the evolution of species and selection by the principle of survival of the fittest. This method generally deals with a population of conformations. Conformations of the next generation are generated by mixing and matching conformations of the current generation. Mutations are also allowed. As the evolution proceeds, those conformations with better fitness (low-energy) survive over those less fit (high-energy conformations). At the end of the algorithm all conformations generally converge into one low-energy conformation (not necessarily the lowest energy conformation). The problem with this approach is that the size of the population must be rather large to avoid the possibility that all conformations converge into a low, but not the lowest energy, local-minimum conformation. Various modifications<sup>3</sup> have been proposed to solve this problem and they have some advantages and disadvantages.

## MCM AND ELECTROSTATICALLY DRIVEN MONTE CARLO

The MCM<sup>5,6</sup> procedure uses the Metropolis MC criterion<sup>7</sup> to search the conformational space of local minima, rather than the whole conformational space. The electrostatically driven MC (EDMC)<sup>8,9</sup> method differs from the MCM procedure in the way that new conformations are generated. It analyzes an accepted conformation based on the orientations of the dipole moments of the protein with respect to the local electric field to generate additional new conformations. This kind of approach is so far the most efficient one for small peptides such as Met-enkephalin, a pentapeptide with the sequence H<sub>2</sub>N-Tyr-Gly-Gly-Phe-Met-COOH, with 24 dihedral angles as variables, shown schematically in Figure 1. Most stochastic methods tend to be trapped in a deep local minimum; it takes a considerably long time to escape from that local minimum in order to sample other regions of conformational space. When this happens in the EDMC method the temperature of the polypeptide under investigation is raised temporarily to a very high degree to alleviate the problem, which is equivalent to trying several random starts.

## CONFORMATIONAL SPACE ANNEALING (CSA)

In this article we propose a method that takes advantage of the buildup procedure and the ge-



**FIGURE 1.** Schematic diagram of the pentapeptide Met-enkephalin. Backbone dihedral angles  $\phi, \psi, \omega$  and side-chain dihedral angles  $\chi^i$  are shown as groups of numbers 1–8. The numbers 1–8 refer to the groups of dihedral angles in the corresponding rectangles in the lower part of the figure. To create trial conformations in the third part of the second step of the algorithm, we substitute dihedral angles of a randomly selected group of the seed conformation with the corresponding dihedral angles of a randomly selected conformation in the bank.

netic algorithm by introducing the concept of CSA. As in the genetic algorithm, we start with a random population of conformations whose energies are then minimized. These local minima are called the “bank.” At the beginning, conformations in the bank are distributed randomly (i.e., minimized from random conformations) in the conformational space of local minima. Conformations in the bank are refined under the condition that they should repel each other in order to cover a large region of conformational space. Schematically, each conformation in the bank is considered as a representative of a group of local minima within a certain distance of each other in the conformational space. Arbitrarily retaining 50 such groups, the following two operations are considered: the first is that each group may improve its representative by replacing the current one with a lower energy conformation within the group. The second is when a new group is found in addition to the existing 50, the group with the highest energy representative among the 51 groups is removed, thereby retaining 50 groups. Therefore, when a new conformation is obtained, it is necessary to determine whether it belongs to one of the 50 existing groups of local minima or not. This is achieved by defining a distance  $D_{ij}$  between two conformations  $i$  and  $j$  in the conformational space and a cutoff value  $D_{\text{cut}}$  that roughly sets the size of each group of local minima in the conformational space by determining whether a new conformation belongs to one of the 50 existing groups or not.

Consider a situation in which improving the conformations in the bank with an additional conformation  $\alpha$  is desired. The generation of conformation  $\alpha$  is described in detail in the second step of the algorithm below. One first finds the conformation  $A$  in the bank that is the closest to the conformation  $\alpha$  at a distance  $D_{\alpha A}$ . If  $D_{\alpha A} < D_{\text{cut}}$ , the conformation  $\alpha$  is considered as one of the local minima belonging to the same group as conformation  $A$ . If conformation  $\alpha$  has a lower energy than conformation  $A$ , then  $\alpha$  becomes the representative of the group and  $A$  is removed from the group without affecting the other groups. However, if  $D_{\alpha A} > D_{\text{cut}}$ , the conformation  $\alpha$  makes up a group itself and the group with the highest energy representative, including the group consisting of conformation  $\alpha$ , is removed, thereby leaving the number of groups unchanged (as 50).

CSA is accomplished by setting the value of  $D_{\text{cut}}$  large enough at the beginning of the algorithm and slowly reducing it to a smaller value. With a large enough value of  $D_{\text{cut}}$ , a small number (e.g., 50) of independent groups of local minima covers the whole conformational space. Therefore, the possibility of discarding a group of local minima with a high-energy representative conformation is eliminated at the beginning of the algorithm because such an initially high-energy conformation may eventually lead to the global minimum-energy conformation. It should also be noted that the lowest energy conformation in the early stages of the optimization does not necessarily lead to the global minimum-energy conformation. The CSA algorithm covers the whole conformational space of local minima at the beginning and slowly replaces groups of high-energy local minima with groups of low-energy local minima as the number of groups of local minima to cover the whole conformational space increases as  $D_{\text{cut}}$  is gradually reduced. However, only a predetermined subset of such groups (e.g., 50) is retained. The value of  $D_{\text{cut}}$  is reduced to a small value (to  $90^\circ$  in 5000 minimizations for the case of Met-enkephalin), and those groups of local minima that fail to obtain representatives with low enough energies relative to other groups are eliminated; that is, only 50 lowest energy groups are retained.

Because the whole conformational space of local minima is effectively divided into many subdivisions by  $D_{\text{cut}}$ , small variations of the conformation in the bank should be sufficient to search for a new representative conformation of the corresponding group. That part of the algorithm in which confor-

mation  $i$  is being improved by varying it is closely related to the buildup procedure in that conformation  $i$  is modified in several stages in which fragments of conformation  $i$  are replaced by corresponding randomly chosen conformations in the rest of the bank:

1. a few random *individual* dihedral angles of conformation  $i$  are replaced by corresponding ones in the bank (for this, use is made of those dihedral angles of the conformations minimized from random starts; see the algorithm in the second section);
2. a randomly chosen *group* of (correlated) dihedral angles; and
3. a randomly chosen *connected group* (comprising three *groups* for Met-enkephalin) of dihedral angles of conformation  $i$  are replaced by corresponding dihedral angles in the bank. (For the definition of the *group* and the *connected group* of dihedral angles see the algorithm in the second section.)

Approximately 3 times as many (correlated) dihedral angles are moved in step 2 as in step 1. Likewise, the ratio of varied dihedral angles in step 3 compared to step 2 is also roughly 3. For peptides larger than Met-enkephalin, additional steps with more dihedral angles moved than in step 3 would be necessary. Steps 2 and 3 are somewhat analogous to setting up two stages of component peptides in the buildup procedure. It should be noted that in the buildup procedure, only a limited number of low-energy conformations of component peptides are stored for use in the next stage. In this new approach, the conformations in the bank are constantly changing. Therefore, although the number of conformations in the bank is limited, any combination of dihedral angles of a group or a connected group is possible in principle.

The proposed method was tested on Met-enkephalin, a model frequently used for evaluating search methods, and the results were compared with those from MCM<sup>5,6</sup> and EDMC.<sup>9</sup> The residue geometry and the interatomic interaction energies used were those in the revised version,<sup>10</sup> ECEPP/3, of the Empirical Conformational Energy Program for Peptides. In ECEPP/3 each conformation of a polypeptide is represented by a set of dihedral angles  $\{\theta_1, \dots, \theta_n\}$ , where  $n$  is the total number of dihedral angles. A simple form of the distance  $D_{ij}$  between two conformations  $i$  and  $j$  in

the conformational space is defined as the sum of the difference of each dihedral angle as follows:

$$D_{ij} = \sum_{k=1}^n \min \left[ \text{mod} \{ (\theta_k^i - \theta_k^j), \text{sym}(k) \}, \right. \\ \left. \{ \text{sym}(k) - \text{mod} \{ (\theta_k^i - \theta_k^j), \text{sym}(k) \} \} \right] \quad (1)$$

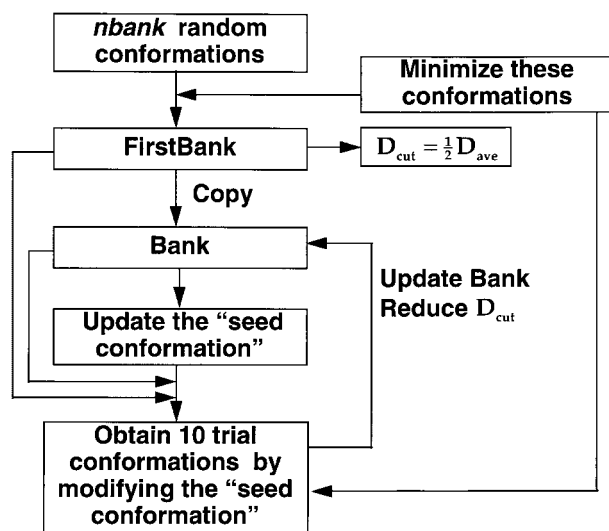
where  $\text{mod}(A, B)$  is the least positive value of  $x$  satisfying  $A = mB + x$  with an integer  $m$ , and  $\text{sym}(k)$  is  $360^\circ$ ,  $180^\circ$ , and  $120^\circ$  for the dihedral angle  $k$  with no symmetry, twofold symmetry, and threefold symmetry, respectively. Thus, distance is defined in terms of dihedral angles. Other forms of distance such as the Euclidean distance of dihedral angles can also be considered.

## Algorithm

A schematic summary of the algorithm is shown in Figure 2 and is described below.

### PREPARATION

We first describe some preparatory procedures before the main algorithm begins.



**FIGURE 2.** Schematic diagram of the algorithm described in the second section. The first four steps are shown schematically in this diagram. The fifth and sixth steps, which are not shown in the figure, are necessary at the end of the round (i.e., when the iteration cannot find a new seed conformation). In such a case, if the number of the round is not more than nine (an arbitrary cutoff) the fifth step is followed, otherwise the sixth step.

Generate *nbank* random conformations (setting all values of  $\omega$  to  $180^\circ$ ), and minimize their ECEPP/3 energies using Gay's secant unconstrained minimization solver, SUMSL<sup>11</sup>; *nbank* = 50 is used for Met-enkephalin with each conformation indexed as *I* (where *I* runs from 1 to 50). Throughout the article the term *minimization* is used to refer to the application of a local minimizer (SUMSL, in particular) to a given conformation. Store the set of minimized conformations as the "first bank." Consider all pairs of distances in this first bank and find  $D_{ave}$ , the average distance between two conformations. Conformations in the first bank, which serves as a reference bank, remain unchanged throughout the rest of the algorithm and provide candidates for local changes of dihedral angles (described in a later section), which are uncorrelated random changes of a few dihedral angles (one to four for Met-enkephalin).

Copy the first bank and save it as the bank. Set the initial value of  $D_{cut}$  as  $D_{cut} = D_{ave}/2$ . Determine the annealing schedule for  $D_{cut}$  as follows. For Met-enkephalin, we have used  $D_{cut} = (D_{ave}/2)x^n$  where *x* is determined so that  $D_{cut}$  becomes  $90^\circ$  after *n* = 5000 minimizations. For *n* > 5000, we used  $D_{cut} = 90^\circ$ . Find the highest energy conformation with index  $I_{max}$  and energy  $E_{max}$ . This conformation is discarded and replaced by a conformation obtained in a later section.

Select the "seed conformation" as the lowest energy conformation in the first bank. It should be noted that the subsequent seed conformation is selected in a different manner from this and is described later. The seed conformation will be modified to obtain new conformations as described in the following subsection.

OBTAIN TRIAL CONFORMATIONS

Here we describe the procedure to obtain trial conformations by modifying the seed conformation. It should be noted that the energies of these trial conformations will be minimized with SUMSL<sup>11</sup>. Information stored in the first bank and the bank will be exploited to modify the seed conformation. We start by defining the dihedral angles around an  $\alpha$ -carbon that are sampled more frequently than the others. These are defined as the dihedral angles  $\phi, \psi$ , and  $\chi^1$  (if there are at least two side-chain dihedral angles). They are involved in rotations around the  $\alpha$ -carbon atom and affect the relative positions of heavy atoms attached to the  $\alpha$ -carbon atom.

Because of steric hindrance, all possible ranges of dihedral angles do not occur uniformly in polypeptides. Therefore, rather than using uniformly distributed random variables for random changes of dihedral angles, we use only those values of the corresponding dihedral angles in the first bank. For example, for a random change of an  $\omega$ , rather than using a value distributed randomly between  $-180^\circ$  and  $180^\circ$ , we use only those values stored in the first bank that are approximately either  $180^\circ$  or  $0^\circ$ . [It should be noted that, although the starting values of all  $\omega$ 's are initially set to  $180^\circ$  (trans conformation), some  $\omega$ 's are close to  $0^\circ$  (cis conformation) after the minimization used to prepare the first bank.] We also classify backbone dihedral angles and side-chain dihedral angles into eight groups, as shown schematically in Figure 1 for Met-enkephalin. In Table I we also show seven possible *connected* groups of three sets of dihedral angles for Met-enkephalin. Groups and connected groups (consisting of three groups) of dihedral angles are analogous to two stages of component peptides in the buildup procedure. Conformations of groups and connected groups are utilized to generate trial conformations (chosen here as 10 for Met-enkephalin) by modifying the seed conformation as follows.

Three conformations are generated by changing a few (one to four for Met-enkephalin) randomly selected dihedral angles of the seed conformation. If the randomly selected dihedral angle of the seed conformation is  $\theta_i$ , we replace it with one of the corresponding  $\theta_i$ 's in the first bank, which is selected randomly out of the *nbank* conformations.

TABLE I.  
List of Connected Groups of Three Sets of Dihedral Angles for Met-Enkephalin, Used for the Second Step of the Algorithm<sup>a</sup>.

Connected Group	Groups
1	1, 2, 3
2	1, 3, 4
3	3, 4, 5
4	4, 5, 6
5	4, 5, 7
6	5, 6, 7
7	5, 7, 8

<sup>a</sup> The specific numbering of each group is given in Figure 1. For example, if the connected group 7 is selected by a random number, the 10 dihedral angles of groups 5, 7, and 8 ( $\phi, \psi, \omega$  of Phenylalanine, and  $\phi, \psi, \omega$  and  $\chi^1, \chi^2, \chi^3, \chi^4$  of Methionine) of the "seed conformation" are replaced by the corresponding dihedral angles of a randomly selected conformation in the bank.

Another three conformations are generated as above but the dihedral angles are selected only from  $\phi$ ,  $\psi$ , and  $\chi^1$ , that is, the dihedral angles around the  $\alpha$ -carbon atom.

Two additional conformations are generated by randomly replacing one of the eight groups of dihedral angles of the seed conformation with a corresponding group of dihedral angles of a randomly selected conformation in the bank.

Two more conformations are generated as above but using connected groups instead of groups. For Met-enkephalin we used seven connected groups comprising three groups of dihedral angles as shown in Table I.

It should be noted that we use the first bank (not the bank) for creating trial conformations in the first two steps. The reason is that dihedral angles in the first bank are used as substitutions for random variables, whereas those in the bank are biased toward specific low-energy conformations because the conformations in the bank are updated in subsequent stages of the algorithm (see the following subsection). Attention is paid to assure that all 10 trial conformations are significantly different (by at least  $30^\circ$  in one dihedral angle) from the seed conformation. We minimize these 10 trial conformations with SUMSL.<sup>11</sup>

### UPDATE BANK AND SELECT NEW SEED CONFORMATION

With the newly obtained minimized conformations, we update the bank as follows. For each newly minimized conformation  $i$  ( $1 \leq i \leq 10$  for Met-enkephalin), find the nearest conformation  $I$  ( $1 \leq I \leq nbank$ ) in the bank with distance  $D_i^{\min}$  from conformation  $i$ .

$$D_i^{\min} = D_{iI} < D_{iJ}, \quad \forall J \neq I \quad \text{and} \quad 1 \leq J \leq nbank. \quad (2)$$

If  $D_i^{\min} < D_{\text{cut}}$ , the new conformation  $i$  is considered to belong to the same group as conformation  $I$  in the bank. Therefore, if  $E_i < E_I$ , conformation  $i$  replaces conformation  $I$ ; otherwise, the bank remains unchanged.

If  $D_i^{\min} > D_{\text{cut}}$ , the conformation  $i$  represents a new group that is distinct from all the other conformations in the bank. We compare  $E_i$  with  $E_{\text{max}}$ , the energy of the highest energy conformation in the bank. If  $E_i > E_{\text{max}}$ , the bank remains unchanged. If  $E_i < E_{\text{max}}$ , conformation  $i$  replaces conformation  $I_{\text{max}}$ , and  $I_{\text{max}}$  and  $E_{\text{max}}$  are updated

appropriately. It should be noted that the number of conformations in the bank remains as  $nbank$ .

Repeat the first three steps above for all newly minimized conformations ( $1 \leq i \leq 10$  for Met-enkephalin) completing the update of the bank. After the bank is updated, a new seed conformation is selected (see next paragraph) and we no longer use the old seed conformation until all bank conformations are used as a seed conformation.

A new seed conformation is selected from the bank as the lowest energy conformation among those satisfying the following two conditions: it has not yet been used as a seed conformation and it lies in the conformational space sufficiently far away (at least by  $d_{\text{ave}}$ ) from the previous seed conformation, where  $d_{\text{ave}}$  is the average distance between the previous seed conformation and all the other conformations satisfying the first condition. The selection of a new seed conformation helps to increase the diversity of conformations in the bank; that is, the new seed conformation is selected significantly far away in the conformational space from the previous one.

### REPEAT SECOND AND THIRD STEPS WITH NEW SEED CONFORMATION

Reduce the value of  $D_{\text{cut}}$  according to the schedule in the first step of the algorithm, and repeat the second and third steps until all bank conformations are used as seed conformations. This completes one round of iterations.

### REPEAT SECOND, THIRD, AND FOURTH STEPS

At the end of each round, all conformations in the bank are designated as unused and a new seed conformation is selected as the lowest energy conformation in the bank. Repeat the second, third, and fourth steps for a preset number of rounds. We set the number of rounds arbitrarily as nine for Met-enkephalin.

### ADD CONFORMATIONS IF NECESSARY

After iterations of the nine preset rounds, we increase the size of the first bank and the bank by adding  $n_{\text{add}}$  more conformations (for Met-enkephalin we used  $n_{\text{add}} = 50$ ). These  $n_{\text{add}}$  conformations are generated as described in the first step of the algorithm. The value of  $D_{\text{cut}}$  is set to the original value that was used in the first step of the algorithm and the second to the fifth steps are

repeated. Special care is needed to prevent the old conformations in the bank from being selected as the seed conformation during the first round after increasing the bank size. This is to minimize the influence of the old conformations with low energies on the evolution of the newly added conformations, the latter having relatively high initial energies because they have not yet been cycled through the optimization procedure.

During the first round of iterations, the number of randomly selected dihedral angles in the first two parts of the second step is set to 2, 3, and 4. After the first round of iterations, it is set to 1.

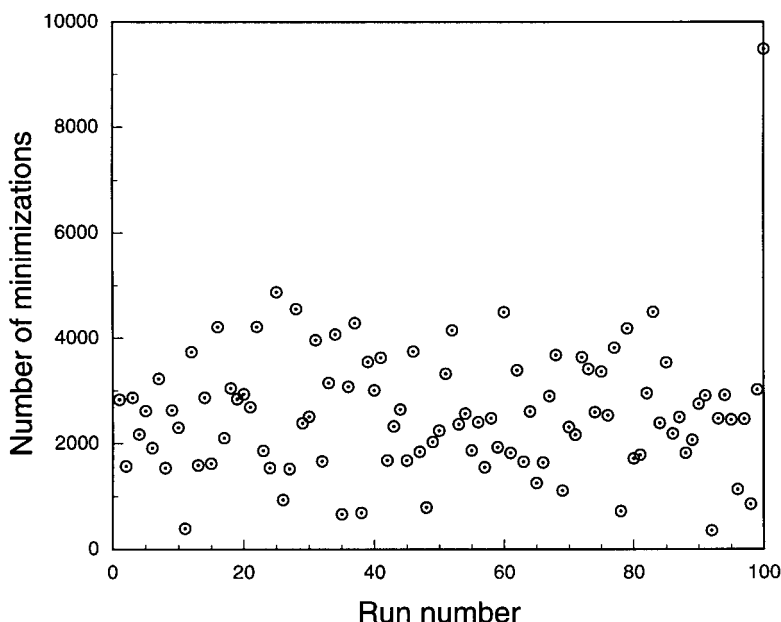
## Results

We applied the CSA method to Met-enkephalin; we tested the algorithm on 100 independent runs and found the same accepted global minimum-energy conformation<sup>10</sup> for all 100 runs. Details of these runs are as follows. The optimization procedure stopped when it found a conformation with  $E < -11.6$  kcal/mol (at this stage the energy was

always  $-11.707$  kcal/mol, which is the accepted value<sup>10</sup> of the ground-state energy).<sup>\*</sup> The initial value of  $n_{\text{bank}}$  was 50, and the number of rounds was set to nine with  $n_{\text{add}} = 50$ . [When necessary, i.e., if the global minimum is not found after nine rounds, then  $n_{\text{add}} = 50$  additional conformations were added to increase  $n_{\text{bank}}$  to 100, which happened only once (the 100th run).] The initial value of  $D_{\text{cut}}$  was set to one-half of  $D_{\text{ave}}$ , and the value of  $D_{\text{cut}}$  was about  $710^\circ$ . We reduced  $D_{\text{cut}}$  algebraically to  $90^\circ$  in 5000 minimizations as described in the first step of the algorithm. Ten trial conformations were generated as described in the second step of the algorithm.

In Figure 3 we show a scatter plot of the number of minimizations to obtain the global minimum-energy conformation for 100 independent runs. For these 100 independent runs, we used one initial value for the random number generator to

<sup>\*</sup>The accepted value reported here differs from the one in ref. 10 because of a redefinition of some of the parameters in the potential function. With the present version of ECEPP/3, we regard  $-11.707$  kcal/mol as the global minimum energy of Met-enkephalin.

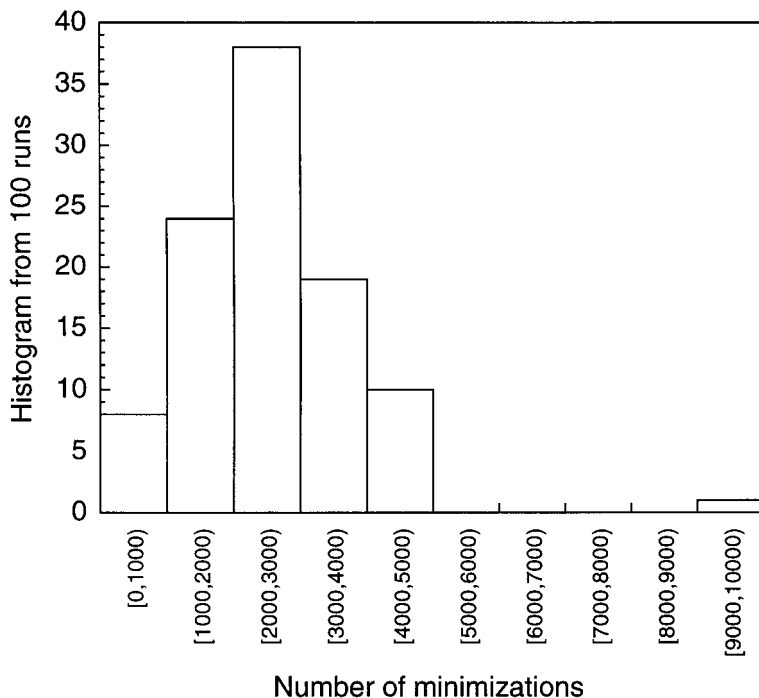


**FIGURE 3.** Scatter plot of the number of minimizations to obtain the accepted global minimum-energy conformation with energy  $E_g = -11.707$  kcal/mol for 100 independent runs. For the 100 independent runs, we used one value for the random number generator to create 100 sets of random first bank conformations of size  $n_{\text{bank}} = 50$ . The global minimum-energy conformation was successfully obtained for all 100 independent runs with about 2600 minimizations on average. It should be noted that the last one was the only one in which the solution was not found within the first nine rounds after 6020 minimizations. The algorithm then added  $n_{\text{add}} = 50$  new conformations before finding the solution within a total of 9482 minimizations.

create 100 sets of random first bank conformations of size  $n_{bank} = 50$ . It should be noted that the 100th run was the only one in which the solution was not found within the first nine rounds after 6020 minimizations, for which the algorithm added  $n_{add} = 50$  new conformations before finding the solution within a total of 9482 minimizations. In the 100th run, the group of local minima containing the global minimum-energy conformation was removed from the bank because it failed in the competition between other groups. That is, the bank conformations that were near the global minimum in conformational space failed to improve their energies after a given number of minimizations and were eventually replaced by other conformations located far from the global minimum as  $D_{cut}$  was reduced from a large to a small value. Figure 4 shows the histogram of the number of minimizations required to obtain the global minimum-energy conformation. On average, it took about 2600 minimizations (corresponding to  $1.7 \times 10^5$  energy evaluations,  $1.3 \times 10^5$  gradient evaluations, and 2700 s of CPU time on a Silicon Graphics Indigo2 workstation), which is an improvement over an average of 3900 minimizations for 16 separate runs from random starts by the EDMC<sup>9</sup>

method. (In ref. 9 all dihedral angles  $\omega$  of the starting conformations were chosen randomly.) In addition, the number of energy evaluations per minimization with the current method is only about 64, which is an improvement over the reported value of about 100 for the MCM<sup>5</sup> method. It seems that this is due to the fact that, in the CSA method, trial conformations are modified only slightly from the seed conformation by substitution of the already minimized portion of the other conformations.

One of the attractive aspects of the proposed method is that a population of distinct local minima is obtained as a by-product, in addition to the global minimum-energy conformation. We examined the final conformations in all 100 runs and collected 1778 conformations with energies lower than  $-9.0$  kcal/mol. It was found that only 194 conformations were unique (obviously, because we found the global minimum in each run, there were 100 identical conformations with energy  $-11.707$  kcal/mol). In Figure 5 the histogram of these 194 conformations are shown. Table II shows the dihedral angles of the 10 lowest energy conformations. The eighth conformation with energy of  $-10.547$  kcal/mol is the most distinctively different confor-



**FIGURE 4.** Histogram of the number of minimizations required to obtain the global minimum-energy conformation. Excluding the 100th, obtained after 9482 minimizations, the distribution is Gaussian-like with a skewness (or third moment)<sup>12</sup> of 0.12 and a kurtosis (or fourth moment)<sup>12</sup> of  $-0.49$ , indicating a slight tail to the right and flatness at the center, respectively, of the distribution.



mation from the global minimum-energy conformation in the table with a distance of  $1259^\circ$ .

## Discussion

The novel aspect of the proposed CSA method is to keep track of many "distinct" local minima that are designed to be distributed as far apart as possible in the conformational space. This is especially important in the early stages of the current method. The landscape of local minima is viewed as a collection of groups of local minima. The physical boundaries between groups of local minima in the conformational space are difficult to define explicitly. In this article they are implicitly defined through the function  $D_{\text{cut}}$  when the bank is updated. The CSA method maintains diversity in the bank by setting the value of  $D_{\text{cut}}$  large enough at the beginning of the algorithm and slowly reducing it, hence conformational space annealing, and by selecting a new seed conformation significantly far away in the conformational space from the previous one.

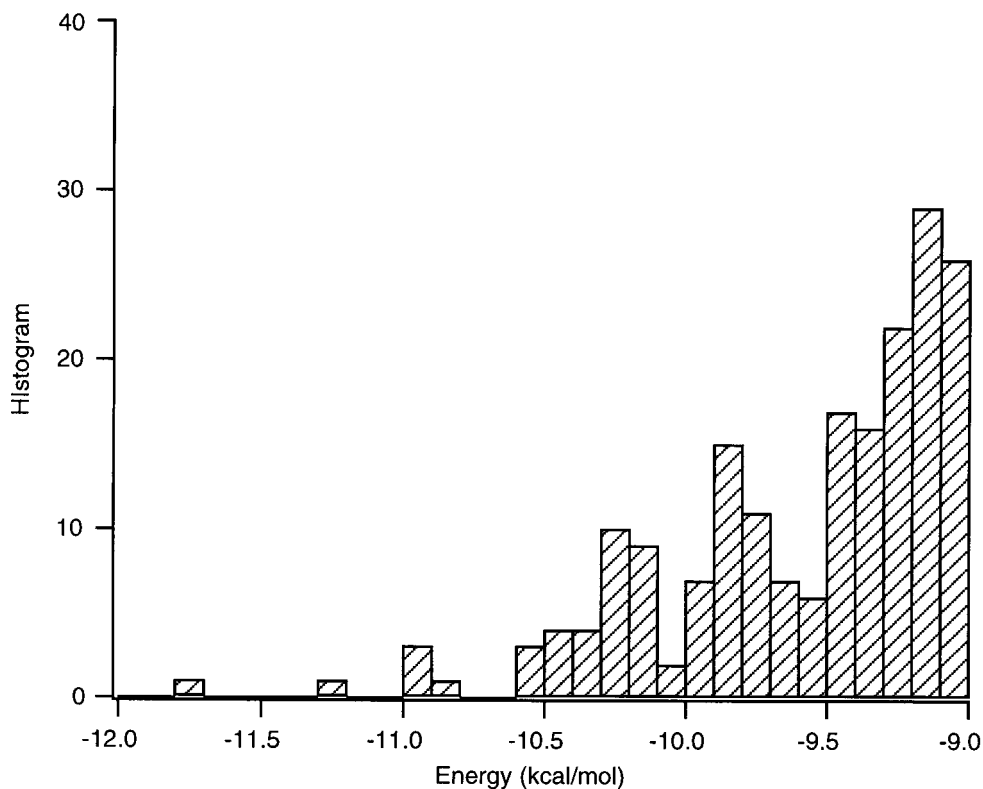
The other important aspects of CSA, borrowed from the genetic algorithm and the buildup procedure, are described in the following subsections.

### ASPECT OF GENETIC ALGORITHM

First, the CSA method deals with many conformations. Second, in obtaining trial conformations, information stored in other conformations is used, which is also related to the buildup procedure.

### ASPECT OF BUILDUP PROCEDURE

For the substitution of a dihedral angle, we use those values from the first bank, the set of minimized conformations from random starts. This is based on the assumption that, if the number of conformations in the first bank is large enough, the global minimum-energy conformation can be obtained by minimizing a conformation made from an appropriate combination of dihedral angles from the first bank. Finding the appropriate combination of dihedral angles is a very difficult optimiza-



**FIGURE 5.** Histogram of 194 low-lying distinct local minima with energies lower than  $-9.0$  kcal/mol. There are 1778 conformations with energies lower than  $-9.0$  kcal/mol from 100 independent runs. Among these 1778 conformations, only 194 conformations are unique; for example, only one of the 194 conformations has an energy between  $-11.7$  and  $-11.8$  kcal/mol, and three different conformations have energies in the range of  $-10.9$  to  $-11.0$  kcal/mol.

**TABLE II.**  
**Dihedral Angles of 10 Lowest-Energy Conformations for Met-Enkephalin<sup>a</sup>.**

		1	2	3	4	5	6	7	8	9	10
Tyr	$\phi$	-83.5	-77.5	58.4	-78.9	-165.9	-78.7	-77.9	-163.2	-155.0	-78.7
	$\psi$	155.8	122.1	154.4	143.0	118.8	141.2	139.4	-40.5	142.9	142.6
	$\omega$	-177.2	-171.1	-176.9	-167.8	-171.0	-168.6	-167.2	-177.7	-167.8	-167.6
	$\chi^1$	-173.2	177.4	-173.9	-164.9	176.1	-168.2	-164.0	-172.2	-165.0	-164.7
	$\chi^2$	79.4	73.8	80.3	75.7	73.8	70.5	79.4	93.2	75.9	76.5
	$\chi^3$	-166.4	-155.1	-166.4	-159.6	-154.0	177.6	-156.1	-177.2	-159.7	-158.7
Gly	$\phi$	-154.3	-104.2	-152.8	-90.1	-100.7	-90.1	-89.8	65.1	-90.0	-90.2
	$\psi$	86.0	61.8	86.3	60.0	63.0	61.0	60.8	-89.7	60.0	60.0
	$\omega$	168.5	167.5	168.4	178.5	167.5	174.5	179.0	174.1	178.4	178.7
Gly	$\phi$	82.9	86.2	83.0	67.3	86.1	74.7	69.2	-152.6	67.3	67.6
	$\psi$	-75.1	-62.5	-75.1	-89.2	-62.8	-77.1	-86.5	34.4	-89.1	-89.0
	$\omega$	-170.0	-173.0	-169.9	175.4	-173.0	-173.7	175.0	-178.9	175.5	175.0
Phe	$\phi$	-136.9	-134.6	-137.0	-94.4	-134.8	-127.5	-100.1	-155.3	-94.5	-95.2
	$\psi$	19.1	18.8	19.2	-31.4	18.8	13.7	-25.5	159.8	-31.4	-31.2
	$\omega$	-174.1	-174.1	-174.0	-171.4	-174.1	-175.5	-172.7	179.2	-171.4	-172.1
	$\chi^1$	58.9	58.5	58.8	179.5	58.5	60.4	72.4	52.1	179.5	-179.3
	$\chi^2$	94.6	94.2	94.6	78.7	94.2	92.3	86.3	82.8	78.7	81.9
	$\chi^3$	-163.5	-162.7	-163.5	-88.7	-162.7	-158.8	-83.3	-79.4	-88.7	-85.5
Met	$\phi$	161.2	157.7	161.3	145.1	157.5	168.3	133.7	130.7	145.0	142.5
	$\psi$	-179.8	-179.5	-179.9	178.9	-179.6	-179.3	178.1	-178.7	179.0	178.8
	$\omega$	52.9	52.5	52.9	-63.8	52.4	53.2	-169.5	-66.7	-63.9	-65.3
	$\chi^1$	175.3	175.5	175.3	-178.3	175.4	173.7	176.2	179.8	-178.3	-71.8
	$\chi^2$	-179.8	-180.0	-179.9	-179.4	-179.9	-179.4	-179.9	-179.9	-179.4	-179.0
	$\chi^4$	-58.6	-58.6	-58.6	-59.9	-58.6	-58.6	60.0	-60.0	-60.0	-59.0
Energy		-11.707	-11.229	-10.980	-10.964	-10.958	-10.857	-10.552	-10.547	-10.531	-10.491
$D_{ii}$		0	175	147	639	258	201	546	1259	705	749

<sup>a</sup> Energies are in kcal / mol and the dihedral angles are in degrees. Dihedral angles of the lowest energy conformation with energy -11.707 kcal / mol agree with those in reference 10 within 3° (see text footnote). Almost all dihedral angles ( $\phi$ ,  $\psi$ ) of conformations 1 and 8 differ from each other by more than 30°. The conformational distances  $D_{ii}$  between conformation 1 and the remaining conformations are calculated from Eq. (1).

tion problem itself. If there are three representative values for each dihedral angle, the number of possible combinations for the 24 dihedral-angle Met-enkephalin becomes  $3^{24} \approx 3 \times 10^{11}$ . At the component-peptides level, we use corresponding groups and connected groups of dihedral angles from the bank to modify the seed conformation. In the buildup procedure, the original problem is built up from single residues to dipeptides, then to oligopeptides, and so on. In the current method, several levels (i.e., a group and a connected group) of modifications of the seed conformation are established simultaneously.

Preliminary results from the CSA method applied to larger polypeptides than Met-enkephalin, such as a collagenlike triple helix,<sup>13</sup> decaglycine, and the 20-residue membrane-bound portion of melittin (not shown here), are very promising. With a local minimizer, the method is quite gen-

eral and it may be applied to other interesting difficult optimization problems.

## Acknowledgment

We thank Y. K. Kang, A. A. Rabow, and D. R. Ripoll for valuable discussions. This work was supported by grants from the National Institute on Aging (AG-00322) and the National Institute of General Medical Sciences (GM-14312) of the National Institutes of Health.

## References

1. M. Vásquez and H. A. Scheraga, *Biopolymers*, **24**, 1437 (1985).
2. K. D. Gibson and H. A. Scheraga, *J. Comput. Chem.*, **8**, 826 (1987).

3. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, AddisonWesley, Reading, MA, 1989.
4. A. A. Rabow and H. A. Scheraga, *Protein Sci.*, **5**, 1800 (1996).
5. Z. Li and H. A. Scheraga, *Proc. Natl. Acad. Sci. USA*, **84**, 6611 (1987).
6. Z. Li and H. A. Scheraga, *J. Mol. Struct. (Theochem.)*, **179**, 333 (1988).
7. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.*, **21**, 1087 (1953).
8. D. R. Ripoll and H. A. Scheraga, *Biopolymers*, **27**, 1283 (1988).
9. D. R. Ripoll and H. A. Scheraga, *J. Protein Chem.*, **8**, 263 (1989).
10. G. Némethy, K. D. Gibson, K. A. Palmer, C. N. Yoon, G. Paterlini, A. Zagari, S. Rumsey, and H. A. Scheraga, *J. Phys. Chem.*, **96**, 6472 (1992).
11. D. M. Gay, *ACM Trans. Math. Software*, **9**, 503 (1983).
12. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes*, Cambridge University Press, New York, 1992, p. 606, eqs. 14.1.5 and 14.1.6, respectively.
13. J. Lee, H. A. Scheraga, and S. Rackovsky, *Peptide Sci.*, to appear.